

CLUSTERING AND VISUALIZATION OF SPECIAL EDITION MAGAZINE SALES ON E-COMMERCE WEBSITES WITH K-MEANS ALGORITHM

Andrew Lomaksan Manuel Tampubolon¹, Thio Marta Elisa Yuridis Butar Butar²,
Siti Rochimah³

Departemen Teknik Informatika, Institut Teknologi Sepuluh Nopember Surabaya

16025231025@student.its.ac.id, 26025231009@student.its.ac.id, 3siti@if.its.ac.id

INTRODUCTION

E-commerce is a digital platform that facilitates online buying and selling transactions between sellers and consumers [1]. The main purpose of e-commerce platforms is to sell products in the form of physical goods and digital services online [2][3]. Sellers on E-commerce platforms are very diverse ranging from individual sellers, small businesses to large-scale companies [1][3]. E-commerce platforms usually offer a variety of features and conveniences such as product catalogs, shopping carts, choices of various payment methods as well as other features that allow customers to find, select, and purchase products or services [3]. In terms of business models, E-commerce can have various types of marketing models, such as online retailers, subscription services, digital marketplaces, and B2B companies that sell products or services to other businesses [3].

Digital technology today is very important for the market and customers so that businesses begin to transform using digital-based media [5]. Some literature has addressed this topic in a fairly segmented manner where studies have considered a single marketing activity or decision and the findings are often discordant [5]. Marketing with both social and digital media has offered significant prospects for entrepreneurs through lower spending, increased brand awareness, and increased sales value [4]. However, a more effective business strategy is needed to be able to maximize existing opportunities [4]. By developing a comprehensive strategy that includes market segmentation and effective resource allocation, businesses can increase their product sales figures and gain a competitive advantage in the market [4].

Artificial intelligence can provide businesses with an understanding of customer characteristics and behavior and help businesses gain a deeper understanding of consumer behavior [6][7][8][9][10]. By analyzing data in the form of sales transactions, artificial intelligence can help businesses identify trends in user activity, gain insights into customer preferences and characteristics, and provide personalized recommendations to improve the customer experience in shopping[7][8]. Analytics with artificial intelligence are important because they enable organizations to make data-driven decisions faster and gain insights into customer characteristics[7][8]. From the customer side, the use of artificial intelligence is very helpful to add deeper consumer insight into a product being marketed and better understand how to categorize and push customers to the next step in their journey[6][7][8][9][10].

The application of artificial intelligence in digital marketing can determine what content is most appropriate and attract customers back to transact [6]. Artificial intelligence offers consumers insights into consumer behavior that are important for customer attraction [7][10]. The application of artificial intelligence to effective sales requires a comprehensive business strategy that includes data strategy, data engineering, governance, change management, and culture [8]. According to Accenture's report, AI: Built to Scale, 84 percent of business executives believe that artificial intelligence can help to achieve their growth goals, but 76 percent admit to struggling with how to improve artificial intelligence across their business [6]. Therefore, it is imperative for businesses to take the necessary steps to develop a comprehensive strategy that includes artificial intelligence to stay ahead and competitive in the market.

In personalizing and staying competitive in the marketplace, artificial intelligence algorithms can be used to segment customers, which is a technique that helps segment customers[7][11][12][13]. A clustering algorithm is an unsupervised machine learning algorithm that finds closely related groups

of data points[7][12]. The k-means clustering algorithm is one example of a common machine learning algorithm suitable for customer segmentation problems [11]. Customer segmentation is the process of classifying customers with similar characteristics into similar segments [7][11][13]. Grouping algorithms help to better understand customer characteristics, both in terms of static demographics and dynamic behavior [7]. By implementing clustering, businesses can group and identify customers into groups that share common product preferences and shopping behaviors. This allows e-commerce businesses to tailor their marketing strategies for each group [14]. It can also increase the level of consumer loyalty and long-term profits [14]. Soft clustering algorithms can be used to achieve a more optimal segmentation of multiple data categories in e-commerce [15]. Some clustering algorithms such as churn can also be used in grouping customers [16]. This can help businesses identify customers who are likely to leave and take proactive measures to retain them.

Zne-Jung Lee et al. proposed a clustering and classification method based on AFE (Automatic Feature Engineering) using a dataset consisting of 4376 customers, 105,170 transaction logs, and 11,171 products. The proposed method has a classification accuracy of 98.49%, which is better than methods such as decision tree, random forest, k-nearest neighbor, and back-propagation network. This method may not be effective if used on datasets that are unstructured or have missing values. In addition, this method also requires computationally intensive purposes, especially when searching for optimal search intervals [17].

Zhao-Hui Sun et al. divided the customer segmentation problem into three stages based on the Customer Requirement Data (CRD) interval. The first stage is data pre-processing which uses entropy to filter out neutral CRDs. The second stage is data transformation. A data transformation scheme based on the standard Gaussian distribution is proposed to model CRDs, which retains the ambiguity of the requirements expressed by the original interval CRDs. The third stage is data grouping. A heuristic grouping method, called GPHC, is proposed. This method uses heuristic information to accurately group CRDs into different classes. Practical case studies are conducted to verify the feasibility and effectiveness of the proposed method. Through proposed solutions (including data pre-processing, data transformation, and GPHC), customer segmentation problems are systematically solved. Companies can directly use the proposed methods to perform customer segmentation tasks, and further adjust market activity in a dynamic customer environment in a timely manner. While this method can be effective in solving customer segmentation problems, there is still plenty of room for further research. For example, how to realize the adaptation of GPHC parameters to avoid manual settings. In addition, as a new heuristic grouping method, popularization and verification of this method in other fields needs to be developed urgently [18].

Jinfeng Zhou et al. present RFMT (Recency, Frequency, Monetary, and Interpurchase Time) models for customer segmentation in the retail industry. This model can be used to identify different groups of customers based on their buying behavior. Segmentation results can provide valuable insights for business strategy, such as allocation of marketing resources and targeted product recommendations. However, there is a lack of information about the methods used to collect customer data. In addition, this paper also does not provide information about the sample size used in the study. This information can be helpful in evaluating the reliability and generalizability of segmentation results. [19]

Yue Li et al. proposed a new customer segmentation method that uses the ALPSO algorithm to improve the accuracy of PSO optimization. This method includes the use of an ALPSO-based K-means algorithm to avoid dependence on the starting point of the cluster. In addition, the IKM-ALPSO method was introduced to handle mixed data in customer segmentation. The experimental results show that this proposed method produces higher accuracy than other methods. This study still lacks execution time for very large datasets. Future research needs to improve and compare the execution time of the IKM-ALPSO clustering algorithm. [20]

Hong-Hao Zhao et al. in this study applying the regulated k-means clustering method with elastic net qualifiers have been introduced to overcome the problem of high-dimensional clustering with correlated variables. The simulation results show that the proposed method produces a lower clustering error rate compared to the standard k-means method. In addition, this method can also select variables simultaneously. In applications on real examples, the number of variables is successfully reduced without losing grouping accuracy. This study has some limitations of only discussing k-means-based grouping methods and does not consider other grouping methods such as EM (Expectation-Maximization). [21]

REFERENCES

1. Rosário, A., & Raimundo, R. (2021). Consumer Marketing Strategy and E-Commerce in the Last Decade: A Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(7), 3003–3024.
2. An Overview of Electronic Commerce (e-Commerce). (2021). *Journal of Contemporary Issues in Business and Government*, 27(3).
3. Chen, Y., Li, M., Song, J., Ma, X., Jiang, Y., Wu, S., & Chen, G. L. (2022). A study of cross-border E-commerce research trends: Based on knowledge mapping and literature analysis. *Frontiers in Psychology*, 13.
4. Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, 59, 102168.
5. Pascucci, F., Savelli, E., & Gistri, G. (2023). How digital technologies reshape marketing: evidence from a qualitative investigation. *Italian Journal of Marketing*.
6. Fauzan M., Davin. (2023). Unlocking Digital Business Success: Leveraging Artificial Intelligence in Social Media Analytics for Enhanced Customer Insights and Engagement. 10.13140/RG.2.2.33588.63361.
7. Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights*, 1(1), 100002.
8. Haleem, A., Javaid, M., Asim Qadri, M., Pratap Singh, R., & Suman, R. (2022). Artificial intelligence (AI) applications for marketing: A literature-based study. *International Journal of Intelligent Networks*, 3, 119–132.
9. Singh, C., Dash, M. K., Sahu, R., & Kumar, A. (2023). Artificial intelligence in customer retention: a bibliometric analysis and future research framework. *Kybernetes*.
10. Jain, V., Wadhvani, K., & Eastman, J. K. (2023). Artificial intelligence consumer behavior: A hybrid review and research agenda. *Journal of Consumer Behaviour*.
11. Siebert, Mark et al. “Technical Background and Methodology for the Elsevier ’ s Artificial Intelligence Report.” (2018).
12. Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160.
13. Sarker, I. H. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, 3(2), 158.
14. Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2018). Customer segmentation in e-commerce: Applications to the cashback business model. *Journal of Business Research*, 88, 407–414.
15. Wu, R.-S., & Chou, P.-H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications*, 10(3), 331–341.
16. Xiahou, X., & Harada, Y. (2022). B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(2), 458–475.
17. Lee, Z.-J., Lee, C.-Y., Chang, L.-Y., & Sano, N. (2021). Clustering and Classification Based on Distributed Automatic Feature Engineering for Customer Segmentation. *Symmetry*, 13(9), 1557.
18. Sun, Z.-H., Zuo, T.-Y., Liang, D., Ming, X., Chen, Z., & Qiu, S. (2021). GPHC: A heuristic clustering method to customer segmentation. *Applied Soft Computing*, 111, 107677.
19. Zhou, J., Wei, J., & Xu, B. (2021). Customer segmentation by web content mining. *Journal of Retailing and Consumer Services*, 61, 102588.
20. Li, Y., Chu, X., Tian, D., Feng, J., & Mu, W. (2021). Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm. *Applied Soft Computing*, 113, 107924.

21. Zhao, H.-H., Luo, X.-C., Ma, R., & Lu, X. (2021). An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation With Correlated Variables. *IEEE Access*, 9, 48405–48412.